

## Reading 14

## Signs, Samples, and Criteria\*

PAUL F. WERNIMONT  
JOHN P. CAMPBELL

Many writers (e.g., Dunnette, 1963; Ghiselli & Haire, 1960; Guion, 1965; Wallace, 1965) have expressed concern about the difficulties encountered in trying to predict job performance, and in establishing the validity of tests for this purpose. In general, their misgivings center around the low validities obtained and misapplications of the so-called "classic validity model." To help ameliorate these difficulties it is proposed here that the concept of validity be altered as it is now applied to predictive and concurrent situations and introduce the notion of "behavioral consistency." By consistency of behavior is meant little more than that familiar bit of conventional wisdom, "The best indicator of future performance is past performance." Surprisingly few data seem to exist to either support or refute this generalization. It deserves considerably more attention.

## SOME HISTORY

It is perhaps not too difficult to trace the steps by which applied psychologists arrived at their present situation. During both World War I and World War II general intelligence and aptitude tests were effectively applied to military personnel problems. Largely as the result of these successes, the techniques developed in the armed services were transported to the industrial situation and applied to the personnel problems of the business organization. From a concentration on global measures of mental ability, validation efforts branched out to include measures of specific aptitudes, interests, and personality dimensions. The process is perhaps most clearly illustrated by the efforts of the United States Employment Service to validate the General Aptitude Test Battery across a wide range of jobs and occupations. In general, testing seemed to be a quick, economical, and easy way of obtaining useful information which removed the necessity for putting an individual on the job and observing his performance over a trial period.

\* P. F. Wernimont and J. P. Campbell, "Signs, Samples, and Criteria," *Journal of Applied Psychology* 52 (1968), 372-76. Copyright 1968 by the American Psychological Association. Reprinted/Adapted by permission of the publisher and author.

It was in the context of the above efforts that an unfortunate marriage occurred, namely, the union of the classic validity model with the use of tests as signs, or indicators, of predispositions to behave in certain ways (Cronbach, 1960, p. 457), rather than as samples of the characteristic behavior of individuals. An all too frequent procedure was to feed as many signs as possible into the classic validity framework in hopes that the model itself would somehow uncover something useful. The argument here is that it will be much more fruitful to focus on meaningful samples of behavior, rather than signs of predispositions, as predictors of later performance.

### THE CONSISTENCY MODEL

To further illustrate the point, consider a hypothetical prediction situation in which the following five measures are available:

1. Scores on a mental ability test;
2. School grade-point average (GPA);
3. Job-performance criterion at Time 1;
4. Job-performance criterion at Time 2;
5. Job-performance criterion at Time 3.

Obviously, a number of prediction opportunities are possible. Test scores could be correlated with GPA; school achievement could be correlated with first-year job success; or the test scores and GPA could be combined in some fashion and the composite used to predict first-, second-, or third-year job performance. All of these correlations would be labeled validity coefficients and all would conform to the classic validity model. It is less clear what label should be attached to the correlation between two different measures of job performance. Few would call it validity; many would probably refer to it as reliability. There seems to be a tendency among applied psychologists to withhold the term validity from correlations between measures of essentially the same behavior, even

if they were obtained at two different points in time. That is, the subtleties of the concept of reliability and the ingredients of the classic validity model seem to have ingrained the notion that validity is a correlation between a predictor and a criterion and the two should somehow be dissimilar.

However, each of the 10 correlations that one could compute from the above situation represents the degree of common variation between the two variables, given the appropriateness of the linear correlation model. After all, that is what correlation is all about. In this sense there is no logical reason for saying that some of the coefficients represent validity and others reliability, although there certainly may be in other contexts. An implicit or explicit insistence on the predictor being "different" seems self-defeating. Rather one should really be trying to obtain measures that are as similar to the criterion or criteria as possible. This notion appears to be at least implicit in much of the work on prediction with biographical data where many of the items represent an attempt to assess previous achievement on similar types of activities. Behavior sampling is also the basis on which simulation exercises are built for use in managerial assessment programs.

At this point it should be emphasized that for the consistency notion to be consistent, the measures to be predicted must also be measures of behavior. For example, it would be something less than consistent to use a behavior sample to predict such criteria as salary progression, organizational level achieved, or subunit production. The individual does not always have substantial control over such variables, and, even with the more obvious biasing influences accounted for, they place a ceiling on the maximum predictive efficiency to be expected. Furthermore, they are several steps removed from actual job behavior. In this respect, the authors are very much in accord with Dunnette (1966) who argues

strongly for the measurement of observable job behavior in terms of its effect on meaningful dimensions of performance effectiveness. A recently developed method for accomplishing this aim is the behavior retranslation technique of Smith and Kendal (1964). The applied psychologist should reaffirm his mandate and return to the measurement of behavior. Only then will one learn by what means, and to what extent, an individual has influenced his rate of promotion, salary increases, or work group's production.

In general terms, what might the selection or prediction procedure look like if one tried to apply a consistency model?

*Process*  
First, a comprehensive study of the job would be made. The results of this effort would be in the form of dimensions of job performance well defined by a broad range of specific behavior incidents which in turn have been scaled with respect to their "criticalness" for effective or ineffective performance.

Next, a thorough search of each applicant's previous work experience and educational history would be carried out to determine if any of the relevant behaviors or outcomes have been required of him or have been exhibited in the past. Items and rating methods would be developed to facilitate judging the frequency of such behaviors, the intensity with which they were manifested, the similarity of their context to the job situation, and the likelihood that they will show up again. These judgments can then be related to similar judgments concerning significant and consistent aspects of an individual's job behavior.

Such a procedure places considerable emphasis on background data and is similar in form to the "selection by objectives" concept of Odiorne and Miller (1966). However, the aim is to be considerably more systematic and to focus on job behavior and not summary "objectives."

After the analysis of background data it

might be found that the required job behaviors have not been a part of the applicant's past repertoire and it would be necessary to look for the likelihood of that job behavior in a variety of work-sample tests or simulation exercises. A number of such behavior measures are already being used in various management assessment programs.

Finally, individual performance measures of psychological variables would be given wider use where appropriate. For example, the Wechsler Adult Intelligence Scale (Wechsler, 1955) might be used to assess certain cognitive functions. Notice that such a measure is a step closer to actual performance sampling than are the usual kinds of group intelligence tests.

How does the above procedure compare to conventional practice? The authors hope they are not beating at a straw man if the usual selection procedure is described as follows. First, a thorough job analysis is made to discover the types of skills and abilities necessary for effective performance. This is similar to the consistency approach except that the objective seems to be a jump very quickly to a generalized statement of skills and abilities rather than remaining on the behavioral level. The conventional approach next entails a search for possible predictors to try out against possible criteria. Based on knowledge of the personnel selection and individual differences literature, personal experience, and "best guesses," some decisions are made concerning what predictors to include in the initial battery. It is the authors' contention that the classic validity model has forced an undue amount of attention on test and inventory measures at this stage. Witness the large amount of space devoted to a discussion of "test validation" in most books dealing with the selection problem. Again, signs seem to take precedence over samples. Lastly, one or more criterion measures are chosen. Too often

the choice seems to be made with little reference to the previous job analysis and is based on a consideration of "objectivity" and relevance to the "ultimate" criterion. Unfortunately, even a slight misuse of these considerations can lead to criteria which are poorly understood. In contrast, working within the framework of a consistency model requires consideration of dimensions of actual job behavior.

It might be added that the above characterization of the conventional approach is meant to be somewhat idealized. Certain departures from the ideal might reinforce the use of signs to an even greater extent. For example, there is always the clear and present danger that the skill requirements will be stated in terms of "traits" (e.g., loyalty, resourcefulness, initiative) and thus lead even more directly to criteria and predictors which are oriented toward underlying predispositions.

## RELATIONSHIP TO OTHER ISSUES

The consistency notion has direct relevance for a number of research issues that appear frequently in the selection and prediction literature. One important implication is that selection research should focus on individuals to a much greater extent than it has. That is, there should be more emphasis on intraindividual consistency of behavior. In their insightful discussion of the criterion problem, Ghiselli and Haire (1960) point out that intraindividual criterion performance sometimes varies appreciably over time, that is, is "dynamic." They give two examples of this phenomenon. However, after an exhaustive review of the literature, Ronan and Prien (1966) concluded that a general answer to the question, "Is job performance reliable?" is not really possible with present data. They go on to say that previous research has not adequately considered the relevant dimensions that contribute to job performance

and very few studies have actually used the same criterion measure to assess performance at two or more points in time. In the absence of much knowledge concerning the stability of relevant job behaviors it seems a bit dangerous to apply the classic validation model and attempt to generalize from a one-time criterion measure to an appreciable time span of job behavior. Utilizing the consistency notion confronts the problem directly and forces a consideration of what job behaviors are recurring contributors to effective performance (and therefore predictable) and which are not.

In addition, the adoption of signs as predictors in the context of the classic model has undoubtedly been a major factor contributing to the lack of longitudinal research. It makes it far too easy to rely on concurrent studies, and an enormous amount of effort has been expended in that direction. Emphasis on behavior samples and behavior consistency requires that a good deal more attention be devoted to the former, along with very explicit consideration of the crucial parameters of a longitudinal study.

The moderator or subgrouping concept also seems an integral part of the consistency approach. The basic research aim is to find subgroups of people in a particular job family for whom behavior on a particular performance dimension is consistent. Subgrouping may be by individual or situational characteristics but the necessity is clear and inescapable. Only within such subgroups is longitudinal prediction possible.

Lastly, the process the authors are advocating demands a great deal in terms of being able to specify the contextual or situational factors that influence performance. It is extremely important to have some knowledge of the stimulus conditions under which the job behavior is emitted such that a more precise comparison to the predictor behavior sample can be made. Because of present difficulties in specifying

the stimulus conditions in an organization (e.g., Sells, 1964), this may be the weakest link in the entire procedure. However, it is also a severe problem for any other prediction scheme, but is usually not made explicit.

It is important to note that the authors' notion of a consistency model does not rest on a simple deterministic philosophy and is not meant to preclude taking account of so-called "emergent" behaviors. Relative to "creativity," for example, the question becomes whether or not the individual has ever exhibited in similar contexts the particular kind of creative behavior under consideration. If a similar context never existed, the research must investigate creative performance and outputs obtained in a test situation which simulates the contextual limitations and requirements in the job situation.

An additional advantage of the consistency approach is that a number of old or persistent problems fortunately appear to dissipate, or at least become significantly diminished. Consider the following:

1. Faking and response sets—Since the emphasis would be on behavior samples and not on self-reports of attitudes, beliefs, and interests, these kinds of response bias would seem to be less of a problem.

2. Discrimination in testing—According to Doppelt and Bennett (1967) two general charges are often leveled at tests as being discriminatory devices:

(a) Lack of relevance—It is charged that test items are often not related to the work required on the job for which the applicant is being considered, and that even where relationships can be shown between test scores and job success there is no need to eliminate low-scoring disadvantaged people since they can be taught the necessary skills and knowledge in a training period after hiring.

(b) Unfairness of content—It is further maintained that most existing tests, espe-

cially verbal measures, emphasize middle-class concepts and information and are, therefore, unfair to those who have not been exposed to middle-class cultural and educational influences. Consequently, the low test scores which are earned are not indicative of the "true" abilities of the disadvantaged. Predictions of job success made from such scores are therefore held to be inaccurate.

The examination of past behaviors similar in nature to desired future behavior, along with their contextual ramifications, plus the added techniques of work samples and simulation devices encompassing desired future behavior, should markedly reduce both the real and imagined severity of problems of unfairness in prediction.

3. Invasion of privacy—The very nature of the consistency approach would seem to almost entirely eliminate this problem. The link between the preemployment or prepromotion behavior and job behavior is direct and obvious for all to see.

## CONCLUDING COMMENTS

The preceding discussion is meant to be critical of the concepts of predictive and concurrent validity. Nothing that has been said here should be construed as an attack on construct validity, although Campbell (1960) has pointed out that reliability and validity are also frequently confused within this concept. Neither do the authors mean to give the impression that a full-scale application of the consistency model would be without difficulty. Using available criteria and signs of assumed underlying determinants within the framework of the classic model is certainly easier; however, for long-term gains and the eventual understanding of job performance, focusing on the measurement of *behavior* would almost certainly pay a higher return on investment.

Some time ago, Goodenough (1949) dichotomized this distinction by referring to

signs versus samples as indicators of future behavior. Between Hull's (1928) early statement of test validities and Ghiselli's (1966) more recent review, almost all research and development efforts have been directed at signs. Relatively small benefits seem to have resulted. In contrast, some recent research efforts directed at samples seem to hold out more promise. The AT&T studies, which used ratings of behavior in simulated exercises (Bray & Grant, 1966), and the In-basket studies reported by Lopez (1965) are successful examples of employing behavior samples with management and administrative personnel. Frederiksen (1966) has reported considerable data contributing to the construct validity of the In-basket. In addition, Ghiselli (1966) has demonstrated that an interview rating based on discussion of specific aspects of an individual's previous work and educational history had reasonably high validity, even under very unfavorable circumstances. In a nonbusiness setting, Gordon (1967) found that a work sample yielded relatively high validities for predicting final selection into the Peace Corps and seemed to be largely independent of the tests that were also included as predictors.

Hopefully, these first few attempts are the beginning of a whole new technology of behavior sampling and measurement, in both real and simulated situations. If this technology can be realized and the consistencies of various relevant behavior dimensions mapped out, the selection literature can cease being apologetic and the prediction of performance will have begun to be understood.

## REFERENCES

- Bray, D. W., & Grant, D. L. The assessment center in the measurement of potential for business management. *Psychological Monographs*, 1966, 80(17, Whole No. 625).
- Campbell, D. T. Recommendations for APA test standards regarding construct, trait, and discriminant validity. *American Psychologist*, 1960, 15, 546-553.
- Cronbach, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper & Row, 1960.
- Doppelt, J. P., & Bennett, G. K. Testing job applicants from disadvantaged groups. *Test Service Bulletin* (No. 57). New York: Psychological Corporation, 1967, pp. 1-5.
- Dunnette, M. D. A modified model for test validation and research. *Journal of Applied Psychology*, 1963, 47, 317-323.
- Dunnette, M. D. *Personnel selection and placement*. Belmont, Calif.: Wadsworth, 1966.
- Frederiksen, N. Validation of a simulation technique. *Organizational Behavior and Human Performance*, 1966, 1, 87-109.
- Ghiselli, E. E. *The validity of occupational aptitude tests*. New York: Wiley, 1966.
- Ghiselli, E. E., & Haire, M. The validation of selection tests in the light of the dynamic character of criteria. *Personnel Psychology*, 1960, 13, 225-231.
- Goodenough, F. *Mental testing: Its history, principles, and applications*. New York: Holt, Rinehart & Winston, 1949.
- Gordon, L. V. Clinical, psychometric, and work sample approaches in the prediction of success in Peace Corps training. *Journal of Applied Psychology*, 1967, 51, 111-119.
- Guion, R. M. Synthetic validity in a small company: A demonstration. *Personnel Psychology*, 1965, 18, 49-65.
- Hull, C. L. *Aptitude testing*. New York: Harcourt, Brace Janovich, 1928.
- Lopez, F. M., Jr. *Evaluating executive decision making: The In-basket technique*. New York: American Management Association, 1965.
- Odierno, G. S., & Miller, E. L. Selection by objectives: A new approach to managerial selection. *Management of Personnel Quarterly*, 1966, 5(3), 2-10.
- Ronan, W. W., & Prien, E. P. *Toward a criterion theory: A review and analysis of research and opinion*. Greensboro, N.C.: Richardson Foundation, 1966.

- Sells, S. B. Toward a taxonomy of organizations. In W. W. Cooper, H. J. Leavitt, & W. W. Shelly, II (Eds.), *New perspectives in organization research*. New York: Wiley, 1964.
- Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating

scales. *Journal of Applied Psychology*, 1963, 47, 149-155.

- Wallace, S. R. Criteria for what? *American Psychologist*, 1965, 20, 411-417.
- Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.

## Reading 15

### Content-Oriented Personnel Selection in a Small Business Setting\*

DAVID D. ROBINSON

A "new emphasis" in the prediction of job behavior was proposed by Wernimont and Campbell (1969). The essence of their idea was that the classic model of criterion-related validity ought to be replaced by a "behavioral consistency" approach to prediction. This approach would rely upon establishment of consistencies between relevant dimensions of job-behavior and preemployment-behavior samples obtained from real or simulated situations.

Guion (1974) pointed out that industrial psychologists paid little attention to content validity until the term was thrust upon them by federal regulations, and concluded that content-referenced measurement constituted a "new window" to be opened. Concepts of job-relatedness and due professional care, emphasized by the courts, e.g., in *Griggs v. Duke Power*<sup>1</sup> and *Albemarle v. Moody*<sup>2</sup> have stimulated interest in content-oriented methodologies.

However, according to Lawshe (1957), the newness of the field and the proprietary nature of the work done by professionals practicing in industry has resulted in a paucity of literature on content validity in employment testing except with regard to the public sector. Prien (1977) has complained that textbooks treat job analysis in such a manner as to "suggest that any fool can do it," and that by doing so relegate it "to the lowest level technician." He asserted the job analysis in test selection and criterion development must not be done by rummaging around in an organization, but through application of highly systematic and precise methods. This paper is offered in response to the apparent needs identified by Lawshe and Prien. Its purpose is to describe a systematic procedure for identifying job content and to illustrate its application to recruiting and selection in a small business setting.

\* D. D. Robinson, "Content-Oriented Personnel Selection in a Small Business Setting," *Personnel Psychology* 34 (1981), pp. 77-87. Copyright 1981 by Personnel Psychology, Inc. Reprinted/Adapted by permission of the publisher and author.

<sup>1</sup> *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).

<sup>2</sup> *Albemarle Paper Co. v. Moody*, 422 U.S. 407 (1975).